PATENT APPLICATION
DOCKET NO. 10008055-1

NETWORK-BASED SPELL CHECKER

INVENTOR(S):

Jerry B. Decime
Marcus Richard Nilson

NETWORK-BASED SPELL CHECKER

FIELD OF THE INVENTION

5   The present disclosure relates to a network-based spell checker.   More particularly, the disclosure relates to a system and method in which a network word search is conducted to help a user to determine the correct spelling of a word.

BACKGROUND OF THE INVENTION

10   Most word processing and electronic mail (email) applications include a spell checking feature, commonly referred to as a "spell checker," that compares words contained within a document with those stored by the application in a word database in an attempt to identify misspelled words.   By way of example, when the spell checker is activated, it scans the document until it identifies an unfamiliar word.   When it does, the

15   spell checker typically generates a pop-up dialogue box that alerts the user that it has located an unknown word which potentially is misspelled.   Typically, the dialogue box also presents several different spelling suggestions to the user.   These suggestions are

selected by the spell checker from the word database according to an algorithm that selects the words that most closely approximate the unfamiliar word located within the document. Due to this configuration, the spell checker may present the user with alternative (and therefore correct) spellings of the intended word and/or other similarly

5 spelled words. Once presented with the suggestions, the user can ignore the suggestions and leave the original spelling, or select one of the suggested words to replace the unfamiliar word.

Conventional spell checkers of the sort described above are limited by their existing "vocabulary," *i.e.*, the collection of words the spell checker maintains in its word

10 database. Accordingly, spell checkers often incorrectly identify correctly spelled, although more recently coined, words. This can be especially true for words that pertain to emerging technologies such as those that support the Internet and the World Wide Web (WWW). For this reason, most spell checkers allow the user to add words that the user presumably believes to be correctly spelled to the word database. For instance, where the

15 spell checker identifies a word it does not recognize which the user believes to be correctly spelled, the user can select an "add" button in the dialogue box to add the word to the database. Once the word is added to the database, the spell checker will recognize it as being correctly spelled next time the spell checker encounters the word.

Although, as noted above, spell checkers normally provide several alternative

20 suggestions to the user, it can be difficult for the user to determine which spelling is correct. If the user is unsure about the correct spelling after viewing the suggestions of

2

the spell checker, the user typically has no choice but to leave the word as originally spelled, guess as to which suggested alternative spelling to select, or consult another reference such as a dictionary to confirm the correct spelling of the word. Clearly, none of these options are very attractive in that the first two may result in the user's document containing a misspelled word and the third is time-consuming or may not even be feasible where the user does not have access to an appropriate reference. Where the user feels relatively sure about the original spelling of a word and decides to add the word to the spell checker word database, the user risks repeating the same spelling mistake over and over if he or she was incorrect as to the spelling of the word.

To alleviate the limitations of conventional spell checkers such as those used in word processing and email programs, Internet-based spell checkers have been created that can be accessed online. Although these spell checkers are more dynamic in that their word databases can be updated by the service provider as new words are coined, limitations to their effectiveness exist. For instance, if a new term is used in a document that has not yet been stored in the word database, the spell checker may incorrectly identify the word as being misspelled. In addition, even where the word is not new, the user may still be unsure about the correct spelling of the word after being presented with the spell checker suggestions. Finally, the utility of any known spell checker, either off line or online, is limited by the various words that it has stored in the word database.

From the foregoing, it can be appreciated that it would be desirable to have a system and method for checking the spelling of words that avoids one or more of the drawbacks identified above.

5

SUMMARY OF THE INVENTION

The present disclosure relates to a method for checking the spelling of words. The method comprises the steps of identifying an unfamiliar word, generating at least one alternative spelling of the unfamiliar word to create a word variant, providing the unfamiliar word and the at least one word variant to a search engine configured to search

10   for a frequency of use of the unfamiliar word and the at least one word variant, and presenting the results of the word search to the user.

The disclosure also relates to a system for checking the spelling of words. The system comprises means for identifying an unfamiliar word, means for generating at least one alternative spelling of the unfamiliar word to create a word variant, means for

15   providing the unfamiliar word and the at least one word variant to a search engine configured to search for a frequency of use of the unfamiliar word and the at least one word variant, and means for presenting the results of the word search to the user.

Furthermore, the disclosure relates to a computer readable medium including a program for checking the spelling of words. The program comprises logic configured to

20   identify an unfamiliar word, logic configured to generate at least one alternative spelling of the unfamiliar word to create a word variant, logic configured to provide the unfamiliar

word and the at least one word variant to a search engine configured to search for a frequency of use of the unfamiliar word and the at least one word variant, and logic configured to present the results of the word search to the user.

Other features, advantages, systems, and methods provided by the invention will become apparent upon reading the following specification, when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, emphasis instead being placed upon clearly illustrating the principles of the present invention.

FIG. 1 is a schematic view of a system for providing spell checking.

FIG. 2 is a schematic view of a computing device shown in FIG. 1.

FIG. 3 is a schematic view of a network server shown in FIG. 1.

FIG. 4 is a flow diagram that illustrates operation of a spell check module shown in FIG. 2.

FIG. 5 is a flow diagram that illustrates operation of a word search engine shown in FIG 3.

DETAILED DESCRIPTION

Referring now in more detail to the drawings, in which like numerals indicate corresponding parts throughout the several views, FIG. 1 illustrates a system 100 for

providing spell checking. As indicated in this figure, the system 100 can comprise one or more computing devices 102 that are each connected to a network 104. As suggested by FIG. 1, the computing devices 102 can have various configurations. For instance, the computing devices 102 can comprise a desktop personal computer (PC) 106 and a handheld

5   device such as a personal digital assistant (PDA) 108. However, as will be apparent from the discussion that follows, the particular configuration of the computing device 102 is important as compared to the fact that the computing device includes an application in which a spell checker can be used and that the computing device is in some way connected to the network 104 (directly or wirelessly) and is therefore capable of communicating with

10   other devices via the network.

The network 104 can comprise one or more sub-networks (*i.e.*, subnets) that are communicatively coupled. By way of example, these networks can include a local area network (LAN) and/or a wide area network (WAN). In a preferred arrangement, however, the network 104 comprises a set of networks that forms part of the Internet. Further

15   included in the system 100 shown in FIG. 1 is at least one network server 110. As indicated in the figure, the server 110 is connected to the network 104, typically through a direct, physical connection.

FIG. 2 is a schematic view illustrating an example architecture for the computing devices 102 shown in FIG. 1. As indicated in FIG. 2, each computing device 102 can

20   comprise a processing device 200, memory 202, one or more user interface devices 204, a display 206, one or more network interface devices 208, and a local interface 210 to

which each of the other components electrically connects. The processing device 200 can include any custom made or commercially available processor, a central processing unit (CPU) or an auxiliary processor among several processors associated with the computing device 102, a semiconductor based microprocessor (in the form of a microchip), or a macroprocessor. The memory 202 can include any one of combination of volatile memory elements (*e.g.*, random access memory (RAM, such as DRAM, SRAM, *etc.*)) and nonvolatile memory elements (*e.g.*, ROM, hard drive, tape, CDROM, *etc.*).

The user interface devices 204 typically comprise those normally used in conjunction with a computing device. For instance, where the computing device 102 comprises a desktop PC, the user interface devices 204 can comprise a keyboard, mouse, *etc.* Where the computing device 102 comprises a handheld device, such as PDA 108, the interface devices 204 can comprise a touch-sensitive liquid crystal display (LCD) and/or one or more function keys. The configuration of the display 206 also normally depends upon the configuration of the computing device 102. For instance, where the computing device 102 comprises a desktop PC, the display typically comprises a monitor. Where the computing device 102 comprises a handheld device, the display 206 can comprise the touch-sensitive screen, where provided, or another LCD provided on the device. The one or more network interface devices 208 comprise the hardware with which the computing device 102 transmits and receives information over the network 104. By way of example, the network interface devices 308 include components that communicate both inputs and outputs, for instance, a modulator/demodulator (*e.g.*,

7

modem), a radio frequency (RF) or other transceiver, a telephonic interface, a bridge, a router, *etc.*

The memory 202 comprises various software and/or firmware programs including an operating system 212, a spell check module 214, network browser 216, and a communications module 218. The operating system 212 controls the execution of other software, such as the spell check module 214, network browser 216, and communications module 218, and provides scheduling, input-output control, file and data management, memory management, and communication control and related services. The spell check module 214 comprises the various software with which, as is described in detail below, the spelling of words in a document can be checked. As used herein, the term "document" refers to any collection of words that contains individual words which can be spell checked. The spell check module 214 uses the network browser 216 to access a network search engine via the communications module 218. The operation of the spell check module 214 is discussed in detail in relation to FIG. 4. Also shown within the memory 202 is a word database 220 that, as is described below, is used to store correctly spelled words that the spell check module 214 can reference when conducting a spell check of a document.

FIG. 3 is a schematic view illustrating an example architecture for the network server 110 shown in FIG. 1. As indicated in FIG. 3, the network server 110 can be similar to that of the computing devices 102 and can therefore comprise a processing device 300, memory 302, one or more user interface devices 304, a display 306, and one or more

network interface devices 308. Each of these components is connected to a local interface 310 that, by way of example, comprises one or more internal buses. The local interface 310 may have additional elements, which are omitted for simplicity, such as controllers, buffers (caches), drivers, repeaters, and receivers to enable communications.

5   Furthermore, the local interface 310 may include address, control, and/or data connections to enable appropriate communications among the aforementioned components.

The processing device 300 comprises hardware for executing software that is stored in memory 302 and can include any custom made or commercially available

10   processor, a central processing unit (CPU), an auxiliary processor among several processors associated with the network server 110, a semiconductor based microprocessor (in the form of a microchip), or a macroprocessor. The memory 302 can include any one of combination of volatile memory elements (*e.g.*, random access memory (RAM, such as DRAM, SRAM, *etc.*)) and nonvolatile memory elements (*e.g.*, ROM, hard drive, tape,

15   CDROM, *etc.*). Moreover, the memory 302 can incorporate electronic, magnetic, optical, and/or other types of storage media. Note that the memory 302 can have a distributed architecture, where various components are situated remote from one another, but accessible by the processing device 300.

The one or more user interface devices 304 typically comprise those normally

20   used in conjunction with a server such as a keyboard, mouse, *etc.*, and the display 306 typically comprises a monitor. The one or more network interface devices 308 comprise

the hardware with which the network server 110 transmits and receives information over the network 104. By way of example, the network interface devices 308 include components that communicate both inputs and outputs, for instance, a modulator/demodulator (*e.g.*, modem), a radio frequency (RF) or other transceiver, a

5   telephonic interface, a bridge, a router, *etc.*

As indicated in FIG. 3, the memory 302 comprises various software programs. In particular, the memory 302 includes an operating system 312, a network search engine 314, and a communications module 316. The operating system 312 controls the execution of other software, such as the network search engine 314 and the

10   communications module 316, and provides scheduling, input-output control, file and data management, memory management, and communication control and related services. The network search engine 314 conducts searches of a database 318 stored within the memory 302 to determine the frequency of use of certain words that are presented to it. This frequency can then be communicated to the user with the communications module

15   316, which operates in conjunction with the network interfaces device(s) 308. The operation of the network search engine 314 is provided below in relation to FIG. 5.

Various software and/or firmware programs have been described herein. It is to be understood that these programs can be stored on any computer readable medium for use by or in connection with any computer related system or method. In the context of this

20   document, a computer readable medium is an electronic, magnetic, optical, or other physical device or means that can contain or store a computer program for use by or in

10

connection with a computer related system or method. These programs can be embodied in any computer-readable medium for use by or in connection with an instruction execution system, apparatus, or device, such as a computer-based system, processor-containing system, or other system that can fetch the instructions from the instruction 5 execution system, apparatus, or device and execute the instructions. In the context of this document, a "computer-readable medium" can be any means that can store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, or device.

The computer readable medium can be, for example but not limited to, an 10 electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, device, or propagation medium. More specific examples (a nonexhaustive list) of the computer-readable medium include an electrical connection having one or more wires, a portable computer diskette, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM, EEPROM, or 15 Flash memory), an optical fiber, and a portable compact disc read-only memory (CDROM). Note that the computer-readable medium could even be paper or another suitable medium upon which a program is printed, as the program can be electronically captured, via for instance optical scanning of the paper or other medium, then compiled, interpreted or otherwise processed in a suitable manner if necessary, and then stored in a 20 computer memory.

FIGs. 4A and 4B illustrate operation of the spell check module 214. As indicated in block 400 of FIG 4A, the spell check module 214 receives a request to check the spelling of the words contained within a document. By way of example, this request can be initiated by the user by, for instance, selecting a check spelling button provided within a word processing application. Alternatively, the request can be initiated by an application, for example email application, automatically in response to some predetermined criterion, *e.g.*, selection of a "send" button of the email application. In any case, the spell check module 214 scans the document for unfamiliar words, as indicated in block 402. In particular, the spell check module 214 searches the word database 220 for each of the words contained within the document to ensure that each is also contained within the word database and, therefore, is correctly spelled.

Flow continues to decision element 404 at which the spell check module 214 determines if an unfamiliar word (*i.e.*, one not contained within the word database 220) is encountered. If not, the words contained within the document are presumably correctly spelled and flow is terminated. If an unfamiliar word is encountered, however, flow continues to block 406 at which the spell check module 214 determines which words to suggest, if any, to the user as a replacement for the unfamiliar word. Typically, the word suggestions are determined according to an algorithm contained within the spell check module 214 that selects one or more correctly spelled words from the word database 220 that are similar to the unfamiliar word. Normally, the algorithm will identify several such words. Once the word suggestions, if any, have been determined, the user is notified that

12

an unfamiliar word has been identified and the suggestions are presented to the user, as indicated in block 408. By way of example, this notification can comprise a message that is presented to the user with a pop-up dialogue box as is conventional in the art.

At this point, it can be determined whether a network word search is requested, as indicated in decision element 410. This request can be generated in a variety of ways. For instance, if the spell check module 214 does not generate any suggestions for the unfamiliar word, the spell check module can be pre-configured to automatically initiate the network word search. In another example, the user can be given the option of requesting the network word search where, for instance, the user is still unsure about the correct spelling of the word even after being presented with the suggestions. If a search request is not generated, flow continues to block 412 at which the user choice received. By way of example, the user can choose a word suggested by the spell check module 214 or opt to ignore the notification and leave the word as originally spelled. Once this selection is made, flow then can return to block 402 where the remainder of the words within the document can be spell checked.

If a network word search is requested, however, flow continues to block 414 in FIG. 4B at which word variants are generated by the spell check module 214. Specifically, the module 214 generates variants of the unfamiliar word that, as is discussed below, will be used as key words in a word search conducted by the network search engine 314 of the server 110. Typically, the word variants are generated by an algorithm of the spell check module 214 that is configured to generate the variants according to certain predetermined

13

rules. By way of example, the algorithm can be adapted to replace vowels and/or consonants of the unfamiliar word with phonetic equivalent (*i.e.*, similar sounding) vowels and/or consonants to generate phonetic equivalent words variants that can be used in the word search. For instance, the algorithm can be configured to replace "u" with "ue" and

5   "ou" with "owe" and so forth until several similar sounding variants of the unfamiliar word are created. Once the variants are created, the network browser 216 can be initiated, as indicated in block 416. Notably, the network browser 216 can be launched by the spell check module 214 immediately once the word search request is received, if desired.

The network browser 216 can comprise an application that is associated with the

10   spell check module 214 and, for instance, provided along with the spell check module in a package arrangement. Alternatively, the spell check module 214 can leverage an existing network browser (*e.g.*, Microsoft™ Internet Explorer™) on the computing device 102. In either case, the network browser 216 accesses the network 104 and, more particularly, accesses one or more network search engines, such as search engine 314 of the network

15   server 110, as indicated in block 418. Like the browser 216, the network search engine 314 can comprise an engine associated with and specifically adapted for the spell check module 214, or can comprise an existing network search engine (*e.g.*, Lycos.com™). Once the search engine is accessed, a word search request can be communicated by the spell check module 214 to the search engine, as indicated in block 420. In particular, the unfamiliar

20   word, as well as several (*e.g.*, ten) of the generated word variants, are provided to the network search engine as key words to be searched for by the engine.

14

At this point, reference is made to FIG. 5 which illustrates the operation of the network search engine 314. As indicated in block 500 of FIG. 5, the network search engine 314 receives the word search request that comprises the unfamiliar word and the generated variants. Once the request, and the various words, have been received, the network search

5 engine 314 conducts a word search of the database 318, as indicated in block 502. By way of example, the database 318 contains a collection of network sites and pages. In a preferred arrangement, the database 318 comprises multitudes of web sites and web pages accessible over the network 104 and, more particularly the World Wide Web. Preferably, the database 318 is frequently updated by the search service provider such that the network

10 search engine 314 can search the most recent documents available which are most likely to contain frequently used words as well as newly coined words that are not yet in common use and therefore unlikely to be recognized by a conventional spell checker. The search engine 314 therefore surveys the database 318 to determine the most frequently appearing, and therefore the most common spellings, of the word at issue.

15 Once the search has been completed, the search results are obtained by the network search engine 314, as indicated in block 504. By way of example, the search engine 314 can be configured to determine the number of "hits" found or percentage of each word. Once this information has been obtained, the search results can be shared with the spell check module 214, as indicated in block 506. Flow for the network search engine 314 is

20 then terminated.

15

Returning to FIG. 4 and block 422, the word search results are received by the spell check module 214. The spell check module 214 then presents these results to the user, as indicated in block 424. More particularly, the searched words and their frequency of use can be presented to the user, and the user can be provided with the option of selecting one of

5   the searched words to replace the word that was unfamiliar to the spell check module 214. These results can be presented in a variety of ways. For instance, the user can be provided with the number of hits for each spelling, the percentage use of each spelling, a graphical representation of frequency, *etc.* Regardless of the particular manner in which the frequency of use information is presented to the user, the user will be able to determine

10   which word or words, and therefore which of the various alternative spellings of the unfamiliar word, is/are most common and, therefore, most likely to be correct. Normally, the correctly spelled word will appear with much greater frequency than the incorrectly "words" such that the correct choice will be clear to the user. Accordingly, the user can normally identify the correct spelling of the word at issue with some degree of certainty.

15   At this point, flow can return to block 412 in FIG. 4A at which the user choice can be received. Again, this choice can be the choice to leave the word as originally spelled or to select one of the word variants that were the subject of the word search conducted by the network search engine 314. Optionally, the user can still choose one of the suggested words originally generated by the spell check module 214 before the word search was conducted.

20   Next, flow returns to block 402 at which the remainder of words contained within the document can be checked for incorrect spelling.

From the above description it can be appreciated that the spell check module 214 can be utilized to not only present word replacement suggestions to the user but also to give the user a better idea as to which of the suggestions is most likely the correctly spelled version of the word. Moreover, it can be appreciated that, due to the network word search,

5   much newer words can be spell checked. It is to be noted that, although the spell check module 214 has been illustrated and described herein as comprising part of the computing device 102, the module could, alternatively, be located in one or more other locations. For instance, persons having ordinary skill in the art will appreciate that the spell check module 214, or a portion thereof, could be stored on a network server, such as server 110, and could

10   be accessed remotely via the network 104. For example, the spell check module 214 could be used in conjunction with an Internet-based email application. Irrespective of its placement, however, the spell check module 214 operates in substantially the same manner to provide the user with greater help in making spelling choices.

While particular embodiments of the invention have been disclosed in detail in the

15   foregoing description and drawings for purposes of example, it will be understood by those skilled in the art that variations and modifications thereof can be made without departing from the scope of the invention as set forth in the following claims.